OPEN ACCESS

KING'S College LONDON

King's Research Portal

DOI:
10.1145/3306446.3340826

Document Version
Publisher's PDF, also known as Version of record

Link to publication record in King's Research Portal

Citation for published version (APA):
Kaffee, L-A., Endris, K. M., & Simperl, E. (2019). When humans and machines collaborate: cross-lingual label editing in Wikidata. In *OpenSym '19 Proceedings of the 15th International Symposium on Open Collaboration* (pp. 1–9). ACM. https://doi.org/10.1145/3306446.3340826

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. May. 2023

When Humans and Machines Collaborate: Cross-lingual Label Editing in Wikidata

Lucie-Aimée Kaffee
kaffee@soton.ac.uk
ECS, University of Southampton, UK
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany

Kemele M. Endris
endris@l3s.de
L3S Research Center
TIB Leibniz Information Centre for Science and Technology
Hannover, Germany

Elena Simperl
E.Simperl@soton.ac.uk
ECS, University of Southampton, UK

ABSTRACT

The quality and maintainability of a knowledge graph are determined by the process in which it is created. There are different approaches to such processes; extraction or conversion of available data in the web (automated extraction of knowledge such as DBpedia from Wikipedia), community-created knowledge graphs, often by a group of experts, and hybrid approaches where humans maintain the knowledge graph alongside bots. We focus in this work on the hybrid approach of human edited knowledge graphs supported by automated tools. In particular, we analyse the editing of natural language data, i.e. labels. Labels are the entry point for humans to understand the information, and therefore need to be carefully maintained. We take a step toward the understanding of collaborative editing of humans and automated tools across languages in a knowledge graph. We use Wikidata as it has a large and active community of humans and bots working together covering over 300 languages. In this work, we analyse the different editor groups and how they interact with the different language data to understand the provenance of the current label data.

KEYWORDS

Multilingual Data, Collaborative Knowledge Graph, Wikidata

ACM Reference Format:
Lucie-Aimée Kaffee, Kemele M. Endris, and Elena Simperl. 2019. When Humans and Machines Collaborate: Cross-lingual Label Editing in Wikidata. In *The 15th International Symposium on Open Collaboration (OpenSym '19), August 20–22, 2019, Skövde, Sweden.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3306446.3340826

This work is licensed under a Creative Commons Attribution International 4.0 License.

OpenSym '19, August 20–22, 2019, Skövde, Sweden
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-6319-8/19/08.
https://doi.org/10.1145/3306446.3340826

1 INTRODUCTION

A wide range of applications today are using linked data, e.g., question answering [2] and natural language generation [1, 7, 8]. Many of those tools are dependent on the natural language representation of the concepts in the knowledge graph. Labels can represent the same concept in a variety of languages. However, the web of data at large lacks of labels in general, multilingual labels in particular [10]. A large international community can help to generate a wider coverage of labels by contributing translations.

Wikidata, a collaborative knowledge graph with a large international community, is widely used in a variety of applications. For instance, it is the structured data backbone of Wikipedia. Wikidata's language distribution is less severe compared to the web at large. However, there is still a strong bias towards English, and the coverage of other languages is lacking [11]. The community of Wikidata consists of humans and bots working alongside each other. This community can contribute to closing the language gap. To understand the provenance of the current label data we analyse the different editor groups and how they contribute to the distribution of languages within labels.

There are different actors contributing to the content of the knowledge graph. We define three groups of editors, analogously to Steiner [18]:

(1) *Registered users*: Editors with an account and a user name. We treat each user name as a different user.
(2) *Anonymous users*: Anonymous users edit without a user account. Instead of a user name, their IP address is recorded. We treat each IP address as one user.
(3) *Bots*: Bots are automated tools that typically work on repeated tasks.

We focus on a comparison of these three different types of editors on a set of different dimensions. We explore the multilinguality among the three user groups, particularly whether automated tools are comparably multilingual to humans, which group is the most active in label editing and

RIGHTSLINK

what kind of patterns we can find in their edit activity over time. We hypothesize that human editors tend to edit in different languages on the same items, i.e. translating labels of one concept, while bots edit different entities in the same language, i.e. importing labels in the same language for a variety of concepts. This would align with the assumption that for a bot one repetitive task (such as importing labels in one language) is easier than a complex task (such as translation of labels in different languages with the context of one item's information). We focus on two editing patterns: (1) a high number of different entities edited and a low number of languages, i.e., monolingual editing over different topics and (2) a low number of different entities and a high number of languages, i.e., translation of labels. Further, we want to understand the connection between languages that editors contribute to.

Finally, we investigate the connection of multilinguality and number of edits. Following the work of [5], who conclude that multilingual editors are more active than their monolingual counterparts, we test whether this holds also for Wikidata editors. The hypothesis is the higher the number of distinct languages per editor, the higher their edit count.

In the following, we first give an overview of the related work in the field of multilingual knowledge graphs and collaboration. Then, we introduce the metrics used in the study to explore the multilingual editing activity of humans and bots in Wikidata. We present and discuss our results in the Sections 4 and 5, and conclude with Section 6.

2 RELATED WORK

Our work focuses on multilingual knowledge graphs. Work in this field has mainly focused on how to construct the ontology (or vocabulary) for such a multilingual knowledge graph [13] or the analysis of the existing content in terms of labels and languages on the web of data [3, 10]. A tool to support the users of a knowledge graph to import labels in other languages is the Labeltranslator introduced by Espinoza et al. [4]. This tool supports the translation of existing labels in other languages.

Collaborative knowledge graphs are edited by a community of users that maintain the knowledge graph in collaboration. Another approach to create and maintain a knowledge graph is by automatic extraction or conversion of data from different sources (e.g. DBpedia [12]). Hybrid approaches, that use automatic tools and human contributions, have the advantage of the large amount of data that can be imported automatically and the precision that human editing has to offer [17]. Our work focuses on Wikidata [20]. Wikidata employs such a hybrid approach, where a community of human editors is supported by automatic tools, so-called bots, that can take over the large amount of mundane and repetitive tasks in the maintenance of the knowledge graph that do not need human decision-making.

We have previously investigated the coverage of multilingual content of Wikidata [11] and conducted a first study on the editing of editors of Wikidata [9]. However, this study is limited to the users that self-assessed their editing languages based on the BabelBox. We extend this work by studying the different user types of Wikidata in-depth and with the background of the difference between humans and bots. We split them into registered editors, anonymous editors, and bots, following the work of Steiner [18]. The author introduces an API for the edit history and conducts analysis based on the data provided. In terms of language edits, they observe which Wikipedia version is most edited by each of the three user groups. Sindhi Wikipedia is purely bot edited, Javanese Wikipedia is purely human edited. They do not apply their metric to Wikidata. Tanon and Kaffee [19] introduce a metric to measure the stability of property labels, i.e. how and whether they change over time. In this work, we use the edit history to draw conclusions on how the labels are edited, similar to our previous work. There have been approaches to explore the editing of multilingual data in Wikidata, particularly of its properties [17], e.g. through visualization [16]. Müller-Birn et al. [14] investigate editing patterns in Wikidata. Wikidata is defined in their work as a combination of factors from the *peer-production* and *collaborative ontology engineering* communities. They also differentiate between human and algorithmic (bot) contributions.

Our understanding of user-contributed content in multiple languages is currently mostly focused on Wikipedia. Wikipedia only covers one language per project, and the different language versions vary widely in size and coverage of topics [6]. In terms of Wikipedia editors, there have been multiple studies on which languages they interact with most. Most relevant for our work is the study on multilingual editors [5]. They introduce a variety of metrics to explore the editing behaviour of Wikipedians across different language Wikipedias. They show that only 15% of editors edit more than one language. However, those multilingual editors make 2.3 more edits than their monolingual counterparts. Park et al. [15] deepen this finding by showing that the edits by multilingual Wikipedia editors are also more complex. Studying the community of Wikidata brings additionally to Wikipedia is an important direction of research, as it is comparable to Wikipedia yet in fundamental parts very different: the project itself is multilingual, the data structure is very different from Wikipedia, there are different shares of editing between bots and humans and there is an overall higher number of daily edits. Bots in Wikidata do repetitive tasks – translation of words is mostly out of scope, however, transliteration or adding names in Latin languages is a feasible task for bots.

RIGHTSLINK

Table 1: Dimensions of the analysis and the metrics applied for each dimension.

Dimension	Metric
User Activity	
	General Statistics
	Edit Timeline
	Edit Count
	Editor Count
Edit Patterns	
	Jumps in Languages
	Jumps in Entities
Language Overlap	
	Connection of languages
	Language family
Activity and Multilinguality	
	Increased Activity

3 METHODS

In this section, we present the dataset and the dimensions used to analyse Wikidata's collaborative editing activity. The code for data preparation and the metrics can be found at https://github.com/luciekaffee/Wikidata-User-Languages/tree/OpenSym2019.

Data preparation

Edit History. Wikidata provides whole dumps of its current data as well as the entire editing history of the project. We worked with a database dump of Wikidata's history, as of 2019-03-01. The data is provided in XML, we converted the data to a PostgreSQL database. The database fields resemble the fields of the XML structure. We extract only label edits by filtering on the *wbsetlabel-set* or *wbsetlabel-add* tag in the edit comment. The history dump includes all information from 2012-10-29 to 2019-03-01. We split the database into three tables (one for each of the user types): registered, anonymous and bots. We define an edit as any alteration of a label, creation, and updating of a label are treated as the same. In the following, we use the term edit only for edits to labels unless specified otherwise.

Users. We split the users into three groups: registered, anonymous and bot editors. Bots on Wikidata are created by community members to import or edit data repetitively and in an automated manner. To ensure that their editing follows the standards of the knowledge graph bots need community approval. Each bot has a unique username and is flagged as a bot. We use the list of bots that have a bot flag on Wikidata[1]. Since historical bots might not currently have a bot flag, we add to the list of bots all users that have a *bot* pre- or suffix, as this is how bots are supposed to be named. Registered users are all users that have a username and do not have a bot flag (or are otherwise marked as bots). Anonymous users do not have a username but an IP address which we treat as a username. This has the disadvantage that we treat each IP address as a single user, not knowing whether the IP address is used by several users. However, this gives us an insight of anonymous users at large, as we can observe their editing patterns in comparison to the other user types.

[1]List of bots with bot flag: https://www.wikidata.org/wiki/Wikidata:Bots

Dimensions

We introduce a set of dimensions, represented as quantitative metrics to measure the multilingual editing activity of different user groups. An overview of all metrics can be found in Table 1.

User Activity. We measure a set of variables related to the activity and multilinguality of the three user groups, that will build the base for the comparison. First, we calculate general statistics: the average number of labels edit per editor, the average number of languages edited per editor, the overall languages covered by each editor type and the average number of editors per language. This gives us a broad insight into the activity of the community. Then, we explore the development of edits over time in the three different groups (*edit timeline*) by summing the edit counts per month. Finally, to understand the support of languages by the editors, we compare edit count and editor count. *Edit count* measures the number of edits per language, and *editor count* measures the number of editors per language. This builds the base to understand the following metrics.

Edit Patterns. We explore the different ways of editing over time between the three different groups. We hypothesize that human editors tend to edit in different languages on the same items, i.e. translating labels of one concept, while bots edit different entities in the same language, i.e. importing labels in the same language for a variety of concepts. We measure these editing patterns by measuring the *jumps* between different languages and entities. For each edit made, we count the number of switches between languages over time. E.g. someone editing (en, en, fr) would have a jump count of 1, i.e., from en to fr, someone editing (fr, de, fr) would have a jump count of 2, i.e., from fr to de and then de to fr. Analogously, we measure jumps between entities. A user editing Berlin's (Q64) label in German and then in French, moving on to the label of the item for London (Q84) in Amharic, i.e. (Q64, Q64, Q84) would have an entity jump count of 1. The numbers are normalized over the total number of edits by user. Generally there are two editing patterns we focus on. First, the part of the community that edits more in one language and therefore has a higher count in jumps of

RIGHTSLINK

Table 2: Results of the general analyses of label editing for the user activity metric. The total number of editors is highest for anonymous editors, their average edit per editor is lowest however. Bots have the lowest number of editors, but the highest number of average edit per editor.

	Registered	Bots	Anon
# Editors	62,091	187	219,127
Avg Edits/Editor	485.2	183,107.6	2.1
Avg Language/Editor	2.2	10.3	1.2
Languages	442	317	369
Avg Editors/Language	310.4	6.13	712.2

entities and lower in languages. Second, the ones that have a higher count in jumps of languages and lower in entities, meaning they translate labels on entities. This metrics can be applied to individual editors in future work. We measure the average over the three different groups to compare them and explore whether there is a tendency differentiating registered users, bots and anonymous user.

Language Overlap. Not only are we interested in the editing behaviour of the community, but also the languages that they edit. We create a language network graph where each node represents a language and the edge represents the cross-lingual edits by a single or more editors. The weight of the edges represents the number of editors that share this language pair. A language pair is the overlap of an editor that edits those two languages. For example, an editor that edits French, German and English creates three connections between those languages (fr-de, de-en, fr-en). Further, we investigate the connection between those language connections and the language families[2] they belong to.

Activity and Multilinguality. We test the hypothesis that a higher number of distinct languages per editor is connected to a higher edit count. We calculate the correlation of those values with Pearson's r, based on the scipy[3] package in Python.

4 RESULTS

We analysed our dataset of label edits based on the metrics introduced in Section 3. We split the dataset into three parts based on the usertype: registered users that edit with a username, anonymous users that edit without a username, and bots, automated tools marked with a bot flag of the *bot* pre- or suffix. In total, we considered 64, 836, 276 edits to labels.

[2]List of language codes and language families: https://github.com/haliaeetus/iso-639/blob/master/data/iso_639-1.json

[3]https://www.scipy.org/

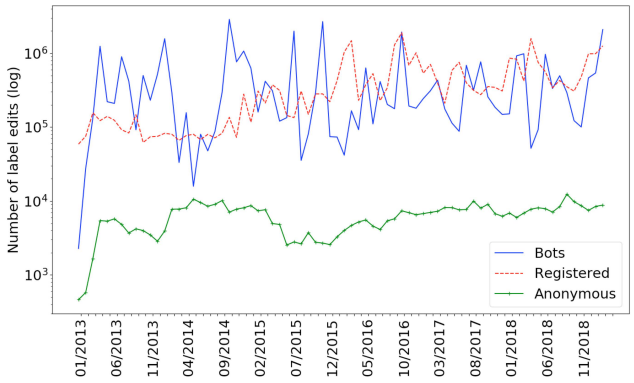


Figure 1: Timeline of number of edits (log) of the three different editor groups from January 2013 to March 2019. Edits are aggregated by month. The highest number of edits for registered users is in October 2016, for bots October 2014 and for anonymous users in September 2018.

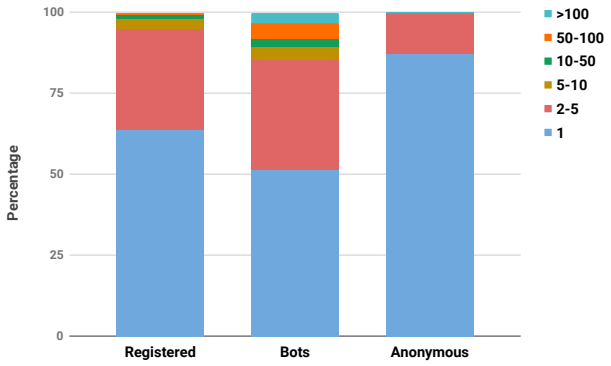


Figure 2: Measuring the distribution of multilingual editors: Each editor type is represented by one bar and split by the number of languages they edit. The majority of editors edit in one language.

Out of all 3, 093, 684 registered users[4], 62,091 users edited labels. This group of editors is responsible for 46.5% of all label edits. The largest group of editors are anonymous editors – a total of 219, 127 unique IP addresses edited Wikidata's labels. However, they contributed to only 0.7% of the label edits. From all bots currently registered with a bot flag[5] and all bots marked with a bot pre- or suffix, 187 bots edited labels. Bots have the highest share of label edits – 52.8% of edits are made by bots.

User Activity. Looking at the average number of edits per editor in Table 2, we find that bots contribute to the

[4]Statistics on users, retrieved March 2019: https://www.wikidata.org/wiki/Special:Statistics

[5]https://www.wikidata.org/wiki/Wikidata:List_of_bots

RIGHTSLINK

Table 3: Bots with the highest numbers of languages edited

Bot name	Languages edited
KLBot2	247
KrBot	240
QuickStatementsBot	150
Cewbot	126
Dexbot	116

Table 4: Average number of *jumps* between languages and edits for all three user groups.

	Registered	Bots	Anon
Languages (Median)	0.2	0.01	0.5
Languages (Avg)	0.3	0.1	0.4
Entities (Median)	0.9	1	0.8
Entities (Avg)	0.8	0.9	0.8

large number of edits not only in total but also on average per bot (183, 107.6). The most active bot (SuccuBot) made 14, 202, 481 edits overall. While there are many anonymous users (219.127), they have a very low edit count per editor (2.1).

For the average number of language per editor, all editor types have a median of 1.0, showing that a majority of editors are monolingual over all three editor types. However, in average registered users and bots have a larger number of languages they edit, showing there are a few very active users compared to the large number of editors editing fewer languages. In Wikipedia, Steiner [18] found that bots are rarely multilingual, showing only ten bots are active in more than five languages. In Wikidata however, bots interact with multiple languages, up to 247 languages (see Table 3). In fact only over half of the bots (51.3%) are monolingual, even less than registered users (63.7%) and anonymous users (87.2, which is to be explained with the low edit count per editor), see Figure 2. Even though registered editors edit fewer languages on average, the multilingual users edit up to 348 languages. Given the small number of edits per editor in the anonymous users, the low number of edits over languages in anonymous users is to be expected.

Figure 3 shows the ranking of languages by edit count and editor count. While the languages overlap neatly for anonymous users (Figure 3c), for the other groups there are strong differences. Given the low edit count by user for anonymous users, the alignment of edit count and editor count is evident. In the other groups, it indicates that more people can edit the language but are less active overall. In all graphs, English is leading for edit count and editor count which aligns with the overall content in Wikidata.

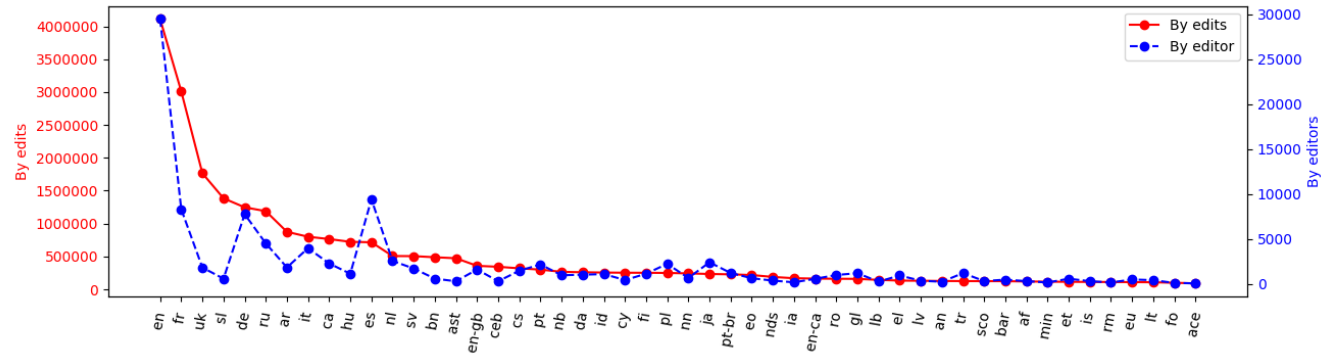
Edit Patterns. We analyse the edit patterns of the different editor types to understand the way the editors edit labels. We measure the change of labels or entities over time in *jumps*. The respective count of jumps is normalized over the total of the edits. We limit this metric to active editors, i.e. editors with at least 500 editors over all time. The results for the normalized numbers of jumps between entities and languages can be found in Table 4. Generally, editors tend to switch more between entities than languages, i.e., there is less translation and more editing of labels in one language over multiple entities. However, there is a slight preference of registered editors to switch between languages compared to bots. Over all their edits, bots tend to edit in one language before switching to the next one.

Language Overlap. We measured the languages that are connected by editors' activity. In Figure 5 we visualize the language connections, limiting them to the ones that are higher than the average, following the work of [5]. For registered users, Figure 5 (a), we see that there is a higher overlap of languages than for bots and anonymous users. While we showed in the previous section and Table 3 that bots edit a variety of languages, the low number of connections in the graph can be explained by the fact that those diverse editing patterns are rare and therefore do not pass the threshold for the weight. Anonymous users have a slightly more diverse editing pattern than bots. However, there are languages connected to only one other node, such as Vietnamese. Those are usually connected to English.

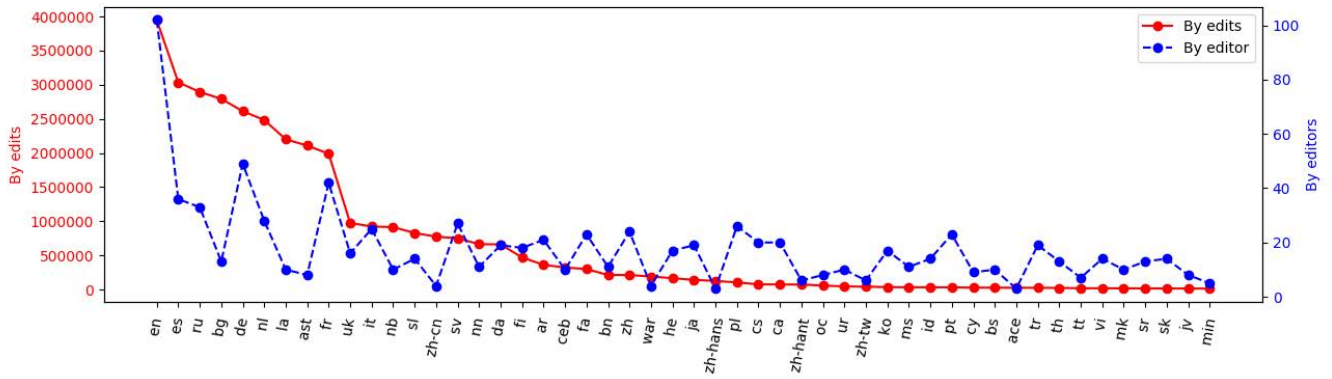
Further, to understand the connection between languages that are edited together and the language families, we counted the number of connections that are in the same language families and compared them to connections in other language families. Figure 4 shows the number of connections for each user group. Even though there is a tendency towards edits in the same language family for all user groups, overall there is no clear connection between language families and editors editing those languages together.

Activity and Multilinguality. We tested the hypothesis that multilingual editors are more active than their counterpart. First, we looked into the percentage of multilingual users, as shown in Figure 2. The majority of users edits in only one language even though even a single edit on a label in a different language would make them *multilingual* in this graph. Figure 6 shows the number of edits (y-axis) and the number of languages edited by the editor (x-axis). There is no clear correlation between the number of languages and the number of label edits as can be seen in the figure. We measured Pearson's r to test the correlation between number of edits and number of languages edited. We used a

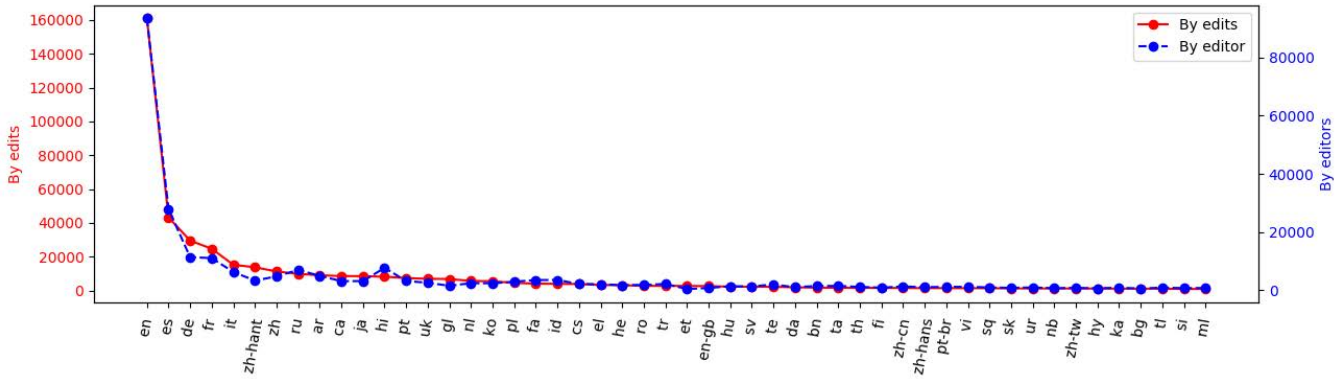
RIGHTSLINK



(a) Registered users



(b) Bots



(c) Anonymous users

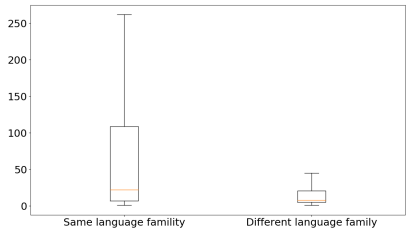
Figure 3: Language Distribution over the three different editor groups, sorted by number of edits, including language ordering by number of editor in that language

two-tailed test. As shown in the previous figure, none of the user groups show a correlation between number of edits and languages (registered editors: (0.21, 0.0), bots: (0.24, 0.001), anonymous: (0.31, 0.0)).

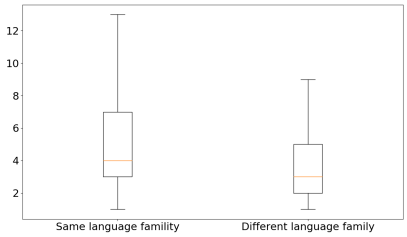
5 DISCUSSION

In this study, we analysed the editing history of Wikidata towards the editing of labels between three user groups: registered editors, bots and anonymous editors. Understanding of label editing is an important topic as labels are the

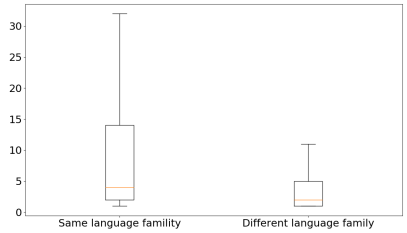
RIGHTSLINK



(a) Registered users



(b) Bots

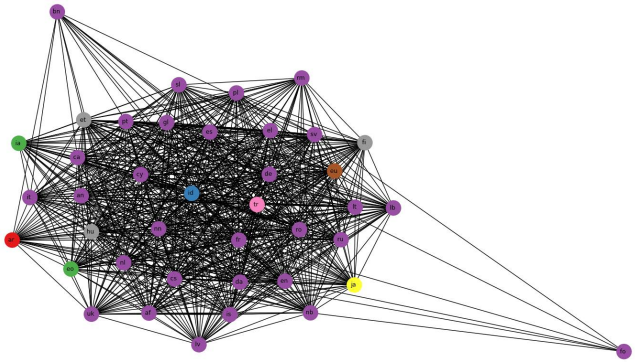


(c) Anonymous users

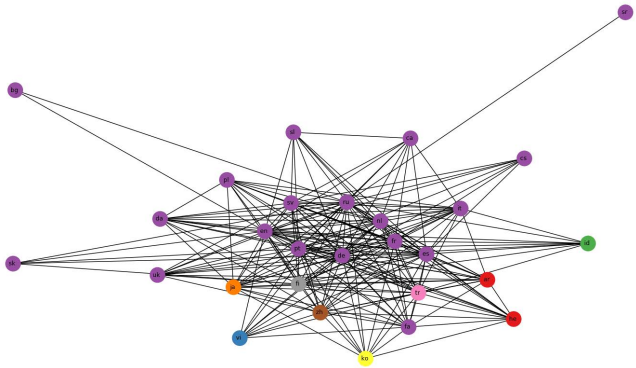
Figure 4: Boxplot comparing the number of edits in languages of the same and different language families.

human-readable representation of the concepts in a knowledge graph. We focus on labels, that are different from other statements in a number of ways: For example, for editors, they can only edit labels if they are somewhat familiar with the language. This work can be extended to other type of statements on a hybrid knowledge graph. We investigate the three user groups towards their label editing and highlight the differences. We find that bots edit by far the most number of edits but less across different languages compared to registered users. Anonymous users have not only low edit count in general and per user, but also less number of edit languages. Active users do not necessarily cover more languages in their editing. Below are the detailed comparisons by user group.

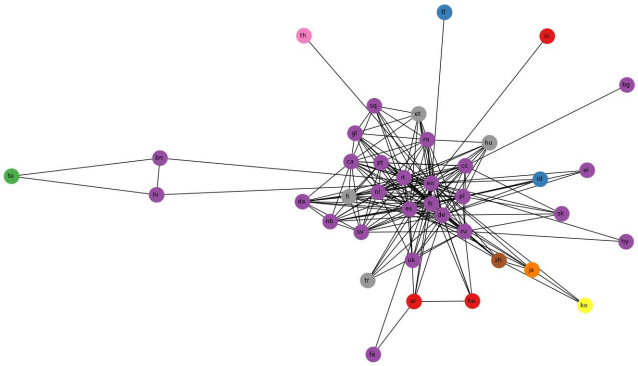
Registered Editors. Registered users accounts the middle between bots and anonymous users: there are less than anonymous users, but they have a higher count of edits per editor. While they edit between languages, they edit fewer languages per editor on average than bots. However, they show a much higher connection between languages than



(a) Registered users



(b) Bots



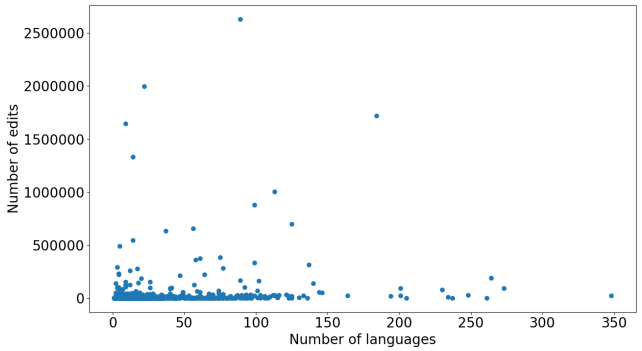
(c) Anonymous users

Figure 5: Displaying the connections between languages, where the number of connections is greater than the average. Nodes are colored by language family.

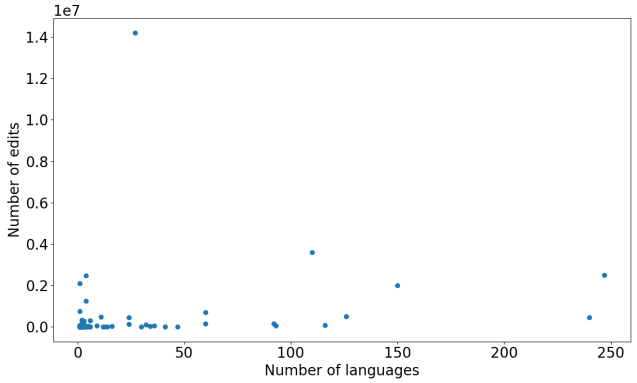
all other user groups. While they are more likely to edit different entities with each edit, they have a higher count of translation (editing different languages after another) than bots.

Bots. Bots, automated tools on Wikidata, have by far the most edit count and contribute to most of the label data even though they are much less in number than registered

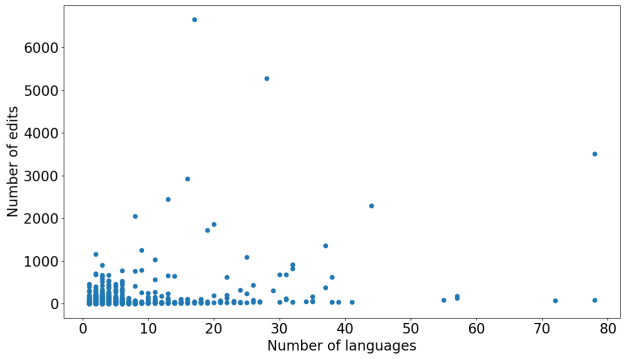
RIGHTSLINK



(a) Registered users



(b) Bots



(c) Anonymous users

Figure 6: Scatter plot of the number of languages and the number of edits, testing correlation for all users.

or anonymous users. A few bots edit a lot of languages, however overall they are not as multilingual as their human counterpart. Compared to bots on Wikipedia, however, they reach much higher counts of languages edited. They are less likely to switch between languages, rather edit one language after another.

Anonymous Editors. Anonymous users are the largest in number but lowest in contribution to label edits. The low number of edits makes it difficult to compare them to the previous groups. Compared to the low edit count per user, there is high cross-lingual activity.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented an analysis of multilingual label editing in Wikidata by three different editor groups. We identify three types of editors: registered editors, bots, and anonymous editors. Bots contributed to the most number of labels for specific languages while registered users tend to contribute more to multilingual labels, i.e., translation.

The hybrid approach of Wikidata, of humans and bots editing the knowledge graph alongside, supports the collaborative work towards the completion of the knowledge graph. The different roles of bots and humans complement each other, as we outline in this work. Future work will deepen the understanding not only of the work that the three editor groups do but also how they interact and support each other and how this can be facilitated. The results of this work can be a starting point for a variety of tools to support the editors, e.g., by suggesting edits to editors based on the knowledge of what bots typically do not do and analogously, suggest the creation of bots for typical bot tasks in labels.

ACKNOWLEDGMENTS

This research was supported by EU H2020 Program for the Project No. 727658 (IASIS); and DSTL under grant number DSTLX-1000094186; and EPSRC-funded project Data Stories under grant agreement number EP/P025676/1.

REFERENCES

[1] Andrew Chisholm, Will Radford, and Ben Hachey. 2017. Learning to generate one-sentence biographies from Wikidata. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, Valencia, Spain, 633–642.

[2] Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret. 2017. Core Techniques of Question Answering Systems over Knowledge Bases: A Survey. *Knowledge and Information systems* (2017), 1–41.

[3] Basil Ell, Denny Vrandečić, and Elena Simperl. 2011. Labels in the Web of Data. *The Semantic Web–ISWC 2011* (2011), 162–176.

[4] Mauricio Espinoza, Asunción Gómez-Pérez, and Eduardo Mena. 2008. Labeltranslator - A Tool to Automatically Localize an Ontology. *The Semantic Web: Research and Applications* (2008), 792–796.

[5] Scott A. Hale. 2014. Multilinguals and Wikipedia Editing. In *ACM Web Science Conference, WebSci '14, Bloomington, IN, USA, June 23-26, 2014*. 99–108. https://doi.org/10.1145/2615569.2615684

[6] Brent J. Hecht and Darren Gergle. 2010. The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15, 2010*. 291–300. https://doi.org/10.1145/1753326.1753370

RIGHTSLINK

[7] Lucie-Aimée Kaffee, Hady ElSahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon S. Hare, and Elena Simperl. 2018. Learning to Generate Wikipedia Summaries for Underserved Languages from Wikidata. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*. 640–645. https://aclanthology.info/papers/N18-2101/n18-2101

[8] Lucie-Aimée Kaffee, Hady ElSahar, Pavlos Vougiouklis, Christophe Gravier, Frédérique Laforest, Jonathon S. Hare, and Elena Simperl. 2018. Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders. In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*. 319–334. https://doi.org/10.1007/978-3-319-93417-4_21

[9] Lucie-Aimée Kaffee and Elena Simperl. 2018. Analysis of Editors' Languages in Wikidata. In *Proceedings of the 14th International Symposium on Open Collaboration, OpenSym 2018, Paris, France, August 22-24, 2018*. 21:1–21:5. https://doi.org/10.1145/3233391.3233965

[10] Lucie-Aimée Kaffee and Elena Simperl. 2018. The Human Face of the Web of Data: A Cross-sectional Study of Labels. In *Proceedings of the 14th International Conference on Semantic Systems, SEMANTICS 2018, Vienna, Austria, September 10-13, 2018*. 66–77. https://doi.org/10.1016/j.procs.2018.09.007

[11] Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. 2017. A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. In *Proceedings of the 13th International Symposium on Open Collaboration*. ACM, 14.

[12] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

[13] Elena Montiel-Ponsoda, Daniel Vila Suero, Boris Villazón-Terrazas, Gordon Dunsire, Elena Escolano Rodríguez, and Asunción Gómez-Pérez. 2011. Style guidelines for naming and labeling ontologies in the multilingual web. (2011).

[14] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production System or Collaborative Ontology Engineering Effort: What is Wikidata?. In *Proceedings of the 11th International Symposium on Open Collaboration, San Francisco, CA, USA, August 19-21, 2015*. 20:1–20:10. https://doi.org/10.1145/2788993.2789836

[15] Sungjoon Park, Suin Kim, Scott Hale, Sooyoung Kim, Jeongmin Byun, and Alice Oh. 2015. MultilingualWikipedia: Editors of Primary Language Contribute to More Complex Articles. In *Ninth International AAAI Conference on Web and Social Media*.

[16] John Samuel. 2018. Analyzing and Visualizing Translation Patterns of Wikidata Properties. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings*. 128–134. https://doi.org/10.1007/978-3-319-98932-7_12

[17] John Samuel. 2018. Towards understanding and improving multilingual collaborative ontology development in Wikidata. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*.

[18] Thomas Steiner. 2014. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata. In *Proceedings of The International Symposium on Open Collaboration, OpenSym 2014, Berlin, Germany, August 27 - 29, 2014*. 25:1–25:7. https://doi.org/10.1145/2641580.2641613

[19] Thomas Pellissier Tanon and Lucie-Aimée Kaffee. 2018. Property Label Stability in Wikidata: Evolution and Convergence of Schemas in Collaborative Knowledge Bases. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*. 1801–1803. https://doi.org/10.1145/3184558.3191643

[20] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM* 57, 10 (Sept. 2014), 78–85. https://doi.org/10.1145/2629489

RIGHTSLINK